Examination and Reconstruction of Three Ancient Endogenous Parvovirus Capsid Protein Gene Remnants Found in Rodent Genomes

Heather M. Callaway,[a] Suriyasri Subramanian,[b,c] Christian A. Urbina,[a] Karen N. Barnard,[a] Robert A. Dick,[d] Carol M. Bator,[b,c] Susan L. Hafenstein,[b,c] Robert J. Gifford,[e] Colin R. Parrish[a]

[a]Baker Institute for Animal Health, Department of Microbiology and Immunology, College of Veterinary Medicine, Cornell University, Ithaca, New York, USA

[b]Department of Medicine, Penn State University College of Medicine, Hershey, Pennsylvania, USA

[c]Department of Biochemistry and Molecular Biology, Penn State University, University Park, Pennsylvania, USA

[d]Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York, USA

[e]MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom

ABSTRACT Parvovirus-derived endogenous viral elements (EVEs) have been found in the genomes of many different animal species, resulting from integration events that may have occurred from more than 50 million years ago to much more recently. Here, we further investigate the properties of autonomous parvovirus EVEs and describe their relationships to contemporary viruses. While we did not find any intact capsid protein open reading frames in the integrated viral sequences, we examined three EVEs that were repaired to form full-length sequences with relatively few changes. These sequences were found in the genomes of *Rattus norvegicus* (brown rat), *Mus spretus* (Algerian mouse), and *Apodemus sylvaticus* (wood mouse). The *R. norvegicus* sequence was not present in the genomes of the closely related species *R. rattus*, *R. tanezumi*, *R. exulans*, and *R. everetti*, indicating that it was less than 2 million years old, and the *M. spretus* and *A. sylvaticus* sequences were not found in the published genomes of other mouse species, also indicating relatively recent insertions. The *M. spretus* VP2 sequence assembled into capsids, which had high thermal stability, bound the sialic acid *N*-acetylneuraminic acid, and entered murine L cells. The 3.89-Å structure of the *M. spretus* virus-like particles (VLPs), determined using cryo-electron microscopy, showed similarities to rodent and porcine parvovirus capsids. The repaired VP2 sequences from *R. norvegicus* and *A. sylvaticus* did not assemble as first prepared, but chimeras combining capsid surface loops from *R. norvegicus* with canine parvovirus assembled, allowing some of that capsid's structures and functions to be examined.

IMPORTANCE Parvovirus endogenous viral elements (EVEs) that have been incorporated into the genomes of different animals represent remnants of the DNA sequences of ancient viruses that infected the ancestors of those animals millions of years ago, but we know little about their properties or how they differ from currently circulating parvoviruses. By expressing the capsid proteins of different parvovirus EVEs that were found integrated into the genomes of three different rodents, we can examine their structures and functions. A VP2 (major capsid protein) EVE sequence from a mouse genome assembled into capsids that had a similar structure and biophysical properties to extant parvoviruses and also bound sialic acids and entered rodent cells. Chimeras formed from combinations of canine parvovirus and portions of the parvovirus sequences from the brown rat genome allowed us to examine the structures and functions of the surface loops of that EVE capsid.

KEYWORDS capsid, endogenous viral element, parvovirus

Endogenous viral elements (EVEs) are generated when viral nucleic acid is incorporated as DNA into the genome of a host organism. EVEs range in size from small gene fragments to whole viral genomes, and they serve as molecular "fossils" if integrated into germ line cell DNA and vertically transmitted from parent to offspring. EVEs have originated from many different families of viruses, with a high proportion being retroviral derived sequences, likely because retroviruses encode reverse transcriptase and integrase proteins that facilitate their integration into host DNA as part of their normal replication cycle. Of the non-retrovirus-derived EVEs, parvoviruses are well represented and likely originate from nonhomologous recombination of the viral DNA genome, possibly facilitated by the single-stranded DNA (ssDNA) nickase activity of the parvovirus replication protein named nonstructural protein 1 (NS1) or replication protein (Rep), which in its normal function creates single-stranded breaks in viral DNA but which can also can nick sequences within the host genome (1–4).

The *Parvoviridae* comprise a family of small ssDNA viruses with genomes of around 5,000 nucleotides. They infect many different animals, ranging from invertebrates to mammals, causing disease in many hosts and likely also causing subclinical infections in many others (5, 6). The different subfamilies of *Parvoviridae* include the parvoviruses (*Parvovirinae*) and the densoviruses (*Densovirinae*) (5). The genomes of the parvoviruses contain two large genes and a variety of smaller open reading frames expressed through alternative splicing or through the use of small transcription units. One large gene encodes the nonstructural protein (NS1 or Rep), which becomes covalently attached to the 5′ end of linear viral ssDNA as well as to double-stranded genome intermediates. Other NS1 functions include acting as a helicase and a site-specific DNA nickase, as well as controlling gene expression from viral and other promoters. Smaller versions of NS1 or Rep are formed through internal initiation of translation or alternative splicing of the viral mRNA, forming proteins that may perform a subset of the functions of the large protein or which provide additional functions such as assisting in viral capsid transport within cells (7–10). The capsid proteins are designated virus protein (VP) or capsid protein (Cap), depending on the virus, and are encoded by a large open reading frame within the right-hand side of the genome. There are a number of capsid protein variants, including a smaller version (VP2 in the autonomous parvoviruses) of 60 to 70 kDa which self-assembles into a stable T=1 capsid (11–13). The largest capsid protein (VP1) contains the complete sequence of VP2 but has an additional sequence (of various lengths between 140 and 250 residues) at the N terminus. VP1 is incorporated into the capsids in various but small amounts (around 10%) in a manner that is structurally equivalent to VP2. The N-terminal unique region of VP1 is packaged within the capsid when first assembled and provides functions required for cell infection, including phospholipase A2 enzyme activity and a nuclear import motif (14, 15). The noncoding regions of the viruses contain promoters for expression of one or more of the viral transcripts, polyadenylation sites, transcriptional regulatory sequences, sequences involved in DNA replication, and the viral genomic termini, which contain terminal repeats that form hairpin or cruciform structures and allow the replication of the viral DNA by host cell DNA polymerases and NS1 or Rep (16–19).

Here, we investigate the properties of some of the many endogenous parvovirus sequences that have integrated into the genomes of different animals. Some of these sequences are ancient, dating to 50 million years old or older based on their orthologous positions in the genomes of divergent species (1, 20–22). In other cases, the sequences have thus far only been found in the genomes of single animal species, suggesting either that they were integrated after the divergence of the last common ancestral animal for which sequences are available or that the integrated copies were deleted from other genomes (21, 22). Parvovirus EVEs most likely derive from nonhomologous recombination of ancient parvoviruses into germ cell genomes, perhaps facilitated by single-stranded breaks created by the nickase function of the NS1 or Rep (1). Retrovirus-derived sequences or transposable elements flank some parvovirus insertions, suggesting that the instability of regions containing transposons or retro-

transposons may also facilitate the insertion of parvovirus or other DNA into host genomes (22). Parvovirus EVEs generally resemble host noncoding DNA or pseudogenes, showing degeneration that is proportional to the length of time they have been integrated, and most do not contain long intact open reading frames due to the presence of base mutations or deletions or insertions of various sizes that result in stop codons or frameshifts within the remaining sequence (20–22). A small minority of parvovirus EVEs, however, contain intact open reading frames, suggesting that some insertions may be of benefit to the host and under selection to maintain the open reading frame (20, 22).

Although the presence of parvovirus-derived EVE sequences in eukaryotic genomes has been recognized for a number of years and more are being found regularly as new genome sequences are obtained, we know little about their biology or whether they play roles in the functions or evolutionary success of their hosts. There has been relatively little analysis of their properties and how they may relate to contemporary viruses or what they tell us about the deeper evolutionary history of the parvoviruses. Here, we examined three parvovirus-derived EVEs that are present in rodent genomes, described their relationships to other members of the parvovirus family, and examined some of their capsid structures and functions. The three VP2-related sequences examined are related to extant viruses and were repaired to give intact open reading frames. The proteins were tested for their ability to assemble into capsids or to be expressed as domains on virus-like particles (VLPs).

RESULTS

Genome mining, examination of hosts related to *Rattus norvegicus*, and reconstitution of parvovirus EVE capsid genes. Whole-genome shotgun (WGS) sequence data of 171 mammalian species (see Data Sets S1 and S2 in the supplemental material) were screened to identify genomic sequences that (i) were derived from ancient viruses sufficiently closely related to modern protoparvoviruses so that they would be amenable to study using these viruses as a model and (ii) encoded complete capsid (VP2) genes. We identified three distinct sequences matching these criteria in the genomes of the brown rat (*Rattus norvegicus*), Algerian mouse (*Mus spretus*), and wood mouse (*Apodemus sylvaticus*). Each occupies a distinct genomic locus and can therefore be assumed to have arisen independently. Critically, phylogenetic reconstructions grouped all three of these sequences together with extant protoparvoviruses in a robustly supported monophyletic clade (Fig. 1A). A few additional sequences disclosing homology to protoparvoviruses were identified in the genomes of rodents, marsupials, and afrotherians (Data Set S2). However, all of these sequences represented only short fragments of VP or were only distantly related to contemporary protoparvoviruses.

The VP2 sequences from all three EVEs contained insertions or deletions that resulted in frameshifts. The *A. sylvaticus* sequence also contained a total of five in-frame stop codons and a stretch of 52 bases that appeared to be missing from the sequence when aligned with the most closely related viable parvoviruses (Fig. 1B to D). To investigate when the *R. norvegicus* EVE integrated during *Rattus* speciation, we used PCR to screen for ~300-bp sequences of the EVE within the DNA of five species in the genus: *R. rattus*, *R. tanezumi*, *R. exulans*, and *R. everetti*, as well *R. norvegicus* as a positive control. We were able to recover sequences only from *R. norvegicus*, putting the age of that EVE insertion at less than 2 million years ago (Mya) (23).

Notably, all three rodent EVEs were more closely related to protoparvoviruses isolated from nonrodent host species than to any that have been found in rodents. Canine parvovirus (CPV) is the extant parvovirus most closely related to the *R. norvegicus* EVE, with their VP2 sequences sharing 68.4% amino acid identity. Most of the differences between CPV and the *R. norvegicus* EVE protein sequences were located on exterior capsid surface loops, while the β-strands and other sequences that form the interior surface of the parvovirus capsid were more conserved (Fig. 2A). The *A. sylvaticus* and *M. spretus* EVEs were both most closely related to porcine parvovirus (PPV), which infects swine, sharing 71.8% and 81.7% amino acid sequence identity with VP2 of that
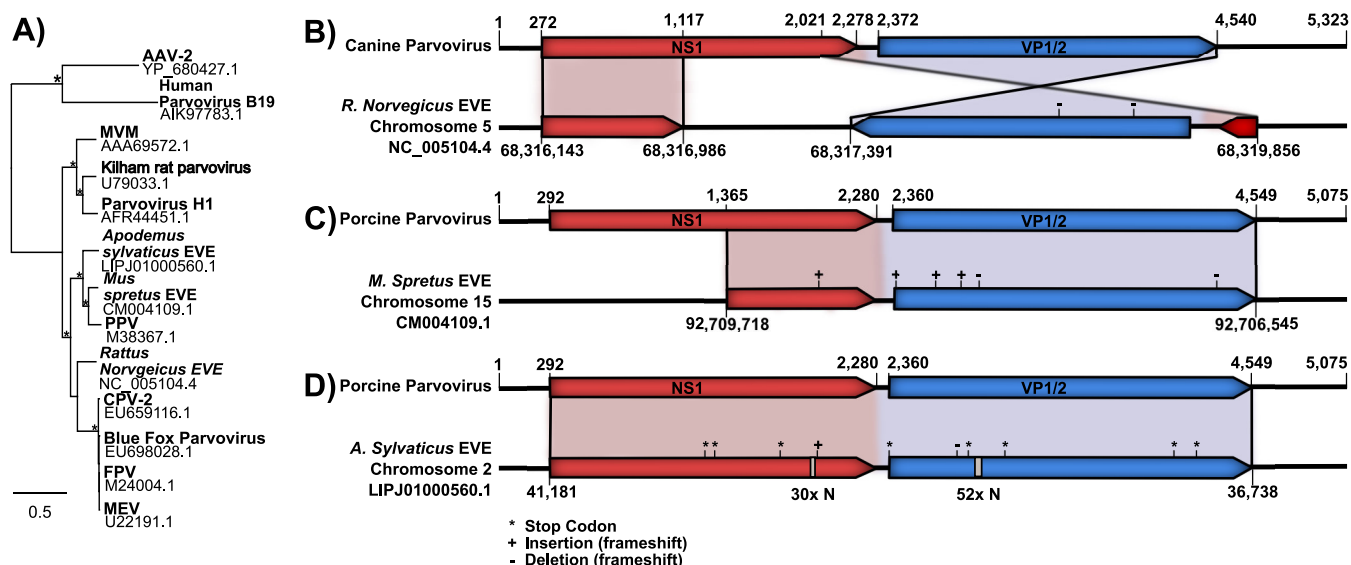
FIG 1 Endogenous viral element insertions in rodent genomes. (A) Phylogenetic tree showing the inferred evolutionary relationships of three rodent EVE sequences and those of different extant protoparvoviruses. Asterisks indicate nodes with bootstrap support ≥90%. The scale bar shows evolutionary distance in substitutions per site. Diagrams of the parvovirus endogenous viral elements used in this study, showing sequence overlap with the most closely related extant virus and indicating the locations of in-frame stop codons and indels that disrupt open reading frames. (B) The *Rattus norvegicus* EVE compared to CPV-2. (C) The *Mus spretus* EVE compared to porcine parvovirus. (D) The *Apodemus sylvaticus* EVE compared to porcine parvovirus.

virus, respectively. Most differences between the *A. sylvaticus* or *M. spretus* EVEs and PPV also occurred in the capsid exterior surface loops, while sequences making up the interior structures of the capsids were more conserved (Fig. 2B and C).

VP2 expression and capsid assembly. After repairing each sequence, the complete VP2 genes were synthesized and inserted into expression plasmids for transfec-
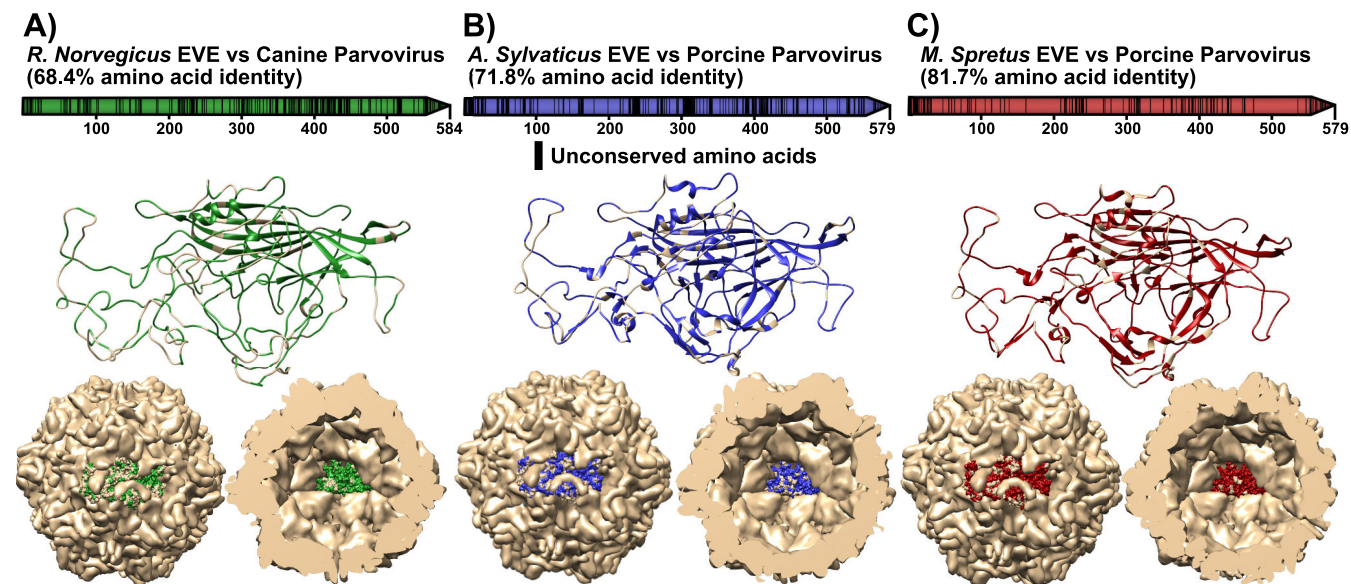

FIG 2 Sequence similarity between the VP2s of the endogenous viral elements and the most closely related extant virus. (A, top) The *R. norvegicus* EVE VP2 sequence, with residues identical to CPV-2 colored green and residues that differ from CPV-2 colored black. Amino acid position is marked. CPV-2 VP2 protein (PDB ID 2CAS) with the capsid exterior facing outward and shared residues with the *R. norvegicus* EVE colored green (middle), shown on the capsid exterior (bottom left) and on the capsid interior (bottom right). (B, top) The *A. sylvaticus* EVE VP2 sequence, with residues identical to PPV colored blue and residues that differ from PPV colored black. Some reference amino acid positions are marked. PPV VP2 protein (PDB ID 1K3V) with the capsid exterior facing outward and shared residues with the *A. sylvaticus* EVE colored blue (middle), shown on the capsid exterior (bottom left) and on the capsid interior (bottom right). (C, top) The *M. spretus* EVE VP2 sequence, with residues identical to PPV colored red and residues that differ from PPV colored black. Some amino acid positions are marked. PPV VP2 protein (PDB ID 1K3V) with the capsid exterior facing outward and shared residues with the *M. spretus* EVE colored red (middle), shown on the capsid exterior (bottom left) and on the capsid interior (bottom right).
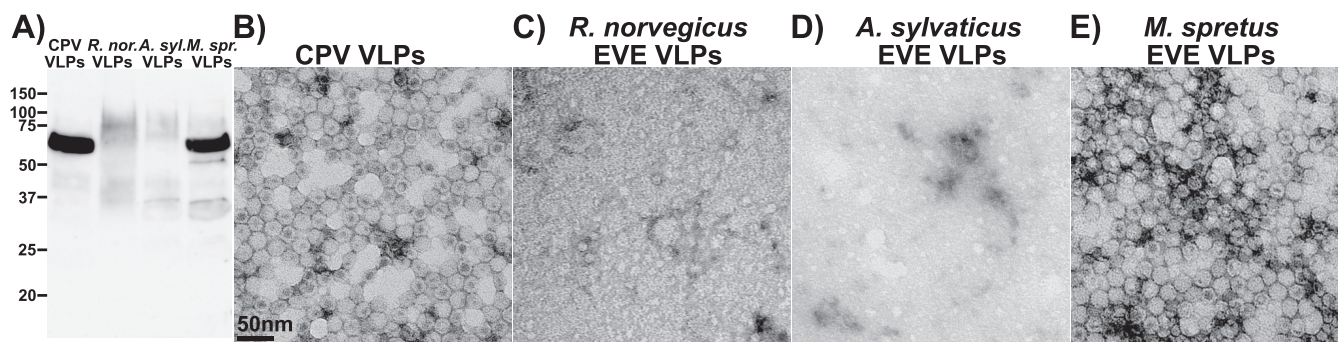
FIG 3 Expression and assembly of *R. norvegicus*, *A. sylvaticus*, and *M. spretus* EVE VP2s. (A) Western blot showing capsid proteins expressed in insect cells and stained with a polyclonal anti-CPV antibody. Negative stain electron microscopy of CPV-2 VLPs (B), *R. norvegicus* EVE VLPs (C), *A. sylvaticus* EVE VLPs (D), and *M. spretus* EVE VLPs (E).

tion into mammalian cells or baculovirus expression in insect cells. When the repaired *R. norvegicus* VP2 sequence was expressed in the HEK293 human-derived cell line, protein expression occurred only at low levels. However, a VP2-like protein was detected in transfected cells by both immunofluorescence assays and Western blots with a rabbit polyclonal antibody prepared against the capsids of feline panleukopenia virus (FPV), which is closely related to CPV. However, electron microscopy of purified samples showed no virus-like particles (VLPs) (data not shown). Baculovirus expression of the *R. norvegicus* VP2 gene produced higher levels of protein expression, but the capsid protein appeared degraded, indicating a lack of assembly (Fig. 3A), and electron microscopy of purified samples from insect cells also showed no VLPs (Fig. 3C).

M. spretus and *A. sylvaticus* VP2 EVE sequences were also expressed in insect cells, and while both EVEs produced capsid proteins that were recognized by a polyclonal antibody produced against SDS-denatured CPV capsids in a Western blot (Fig. 3A), only the *M. spretus* EVE VP2 assembled into capsids (Fig. 3D and E).

***R. norvegicus* EVE/CPV capsid sequence chimeras.** We created chimeric VP2s consisting of the CPV VP2 gene with segments replaced by equivalent sequences from the *R. norvegicus* EVE. Initial chimeras used restriction enzyme sites shared between the CPV and *R. norvegicus* EVE sequences to divide VP2 approximately into thirds, and those sequences were ligated together in all 6 possible combinations and expressed in insect cells (Fig. 4). While each of the chimeras expressed protein (Fig. 4A), most showed altered molecular weight products, suggestive of protein expression but degradation. Negative stain electron microscopy of purified fractions also revealed no capsids (Fig. 4C to H).

A second set of chimeras had the individual exterior surface loops from the CPV-2 replaced with the corresponding sequence of the *R. norvegicus* EVE (Fig. 5A and B). In the parvovirus capsid structure, there are 4 loops inserted between the β-strands of the β-barrel or jelly roll structure. Those are labeled loops B-C (loop 1), E-F (loop 2), the first half of the G-H loop (loop 3), the second part of the G-H loop (loop 4), as well as the end of loop I to the C terminus, which we labeled loop 5 (24) (Fig. 5A and B). Chimeras with individual surface loop substitutions were expressed in insect cells, and capsids were purified where present and examined using negative stain electron microscopy (Fig. 5C and D). Chimeras containing each of the *R. norvegicus* EVE surface loops 2, 3, 4, or 5 individually assembled into capsids, but assembly was not seen for those containing surface loop 1 (Fig. 5D). A chimera containing all five surface loops also failed to assemble into VLPs (data not shown), likely because it contained surface loop 1.

We purified VLPs of chimeras containing individual *R. norvegicus* surface loops 2, 3, 4, or 5, incubated them with either Norden laboratory feline kidney (NLFK) cells (which support CPV infection) or murine L cells, and measured uptake. The NLFK and L cells were chosen as being similar to the feline and rodent hosts that the different viruses
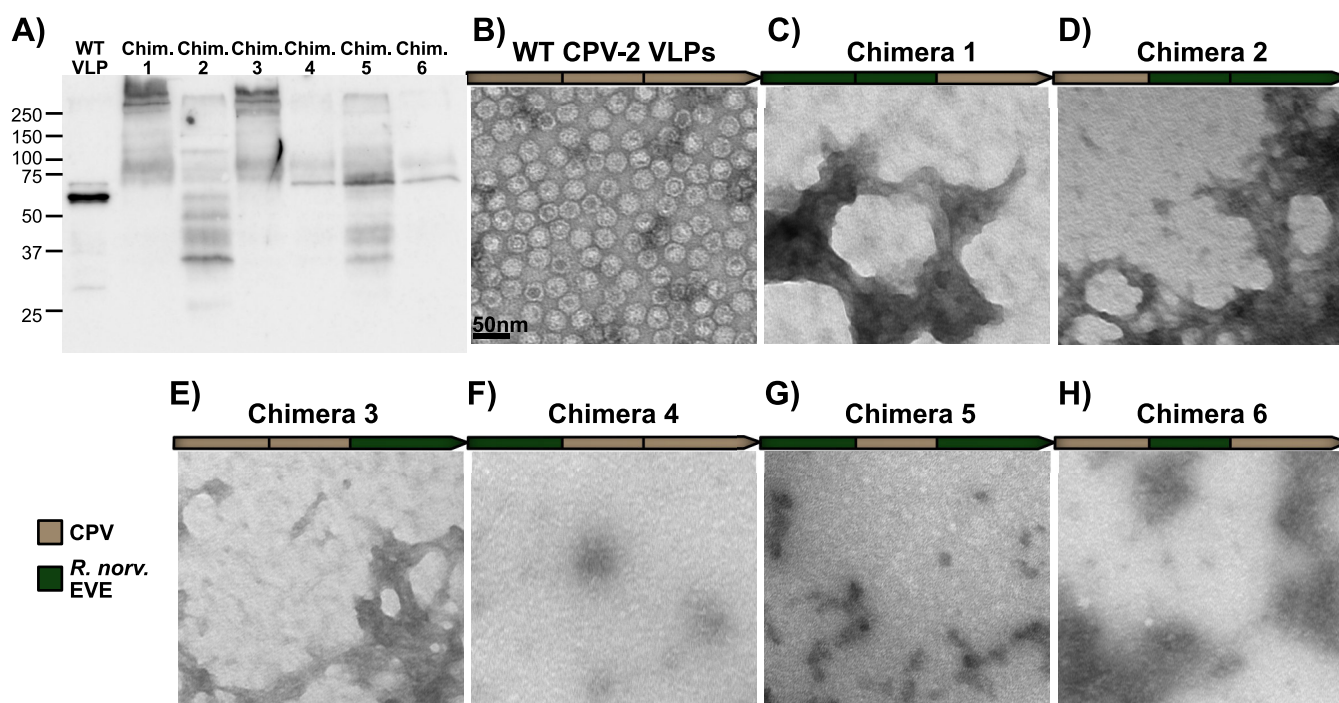
FIG 4 Expression and assembly of chimeras constructed by combining BglII and PhoI restriction enzyme digestions of the *R. norvegicus* EVE with those of CPV-2. (A) Western blot showing capsid proteins chimeras or CPV-2 expressed in insect cells and stained with a polyclonal anti-CPV antibody. (B) Electron micrograph of CPV-2 VLPs showing assembled capsids. (C to H) Electron micrographs of chimeras between the *R. norvegicus* EVE and CPV-2, showing that capsid proteins do not assemble. Diagrams of the chimeras are also shown above each micrograph, with the VP2 gene divided at the BglII and PhoI restriction enzyme sites and the *R. norvegicus* EVE represented in green.

in the chimeras were associated with. The loop 2, 3, or 4 chimeras each replaced a CPV surface loop associated with binding to the transferrin receptor type-1 (TfR), the natural receptor for CPV capsids (25). These chimeras either did not enter NLFK cells (loop 2 and 4 chimeras) or displayed an altered pattern of uptake, where VLPs did not colocalize with the TfR (loop 3 chimeras) (Fig. 6A). Loop 5 chimeric capsids showed the same pattern of uptake by NLFK cells as CPV capsids, suggesting that they bind to the TfR (Fig. 6A). While CPV VLPs or loop 2, 4, or 5 chimeras did not bind or enter L cells, loop 3 chimeras did (Fig. 6B), indicating that the loop contains the receptor binding site for the *R. norvegicus* EVE.

M. spretus EVE capsid structure and function. The *M. spretus* VLPs were very stable, with an average melting temperature of 82°C when tested for temperature stability by differential scanning fluorimetry (DSF) (Fig. 7A and B). While changing the pH from 7.4 to 5.5 increased the melting temperature of CPV-2 VLPs by approximately 6°C, that of the *M. spretus* VLPs was not affected by the altered pH. The derivative curves from *M. spretus* VLPs did, however, show a minor peak that was affected by changes in pH (Fig. 7A), which may represent a transition in that capsid structure prior to full disassembly. The *M. spretus* VLPs also formed fewer low-molecular-weight proteolytic cleavage products than CPV-2 VLP controls (Fig. 7C and D).

Fluorescently labeled *M. spretus* capsids bound and entered murine L cells and accumulated in a perinuclear location, indicating endocytosis, and were also able to enter feline NLFK cells (Fig. 7E). *M. spretus* capsids hemagglutinated mouse erythrocytes (Fig. 7F), which contained over 97% *N*-acetylneuraminic acid (Neu5Ac) (Table 1). Neuraminidase treatment of those erythrocytes eliminated hemagglutination by the capsids (Fig. 7F), showing that the binding was to sialic acids.

We determined the structure of the *M. spretus* EVE VLP at a resolution of 3.89 Å using cryo-electron microscopy (cryo-EM) (Fig. 8). Superimposing that structure with the capsid structures of minute virus of mice (MVM) strains MVMi and MVMp, porcine parvovirus (PPV), or CPV showed that the *M. spretus* capsid resembled those of other
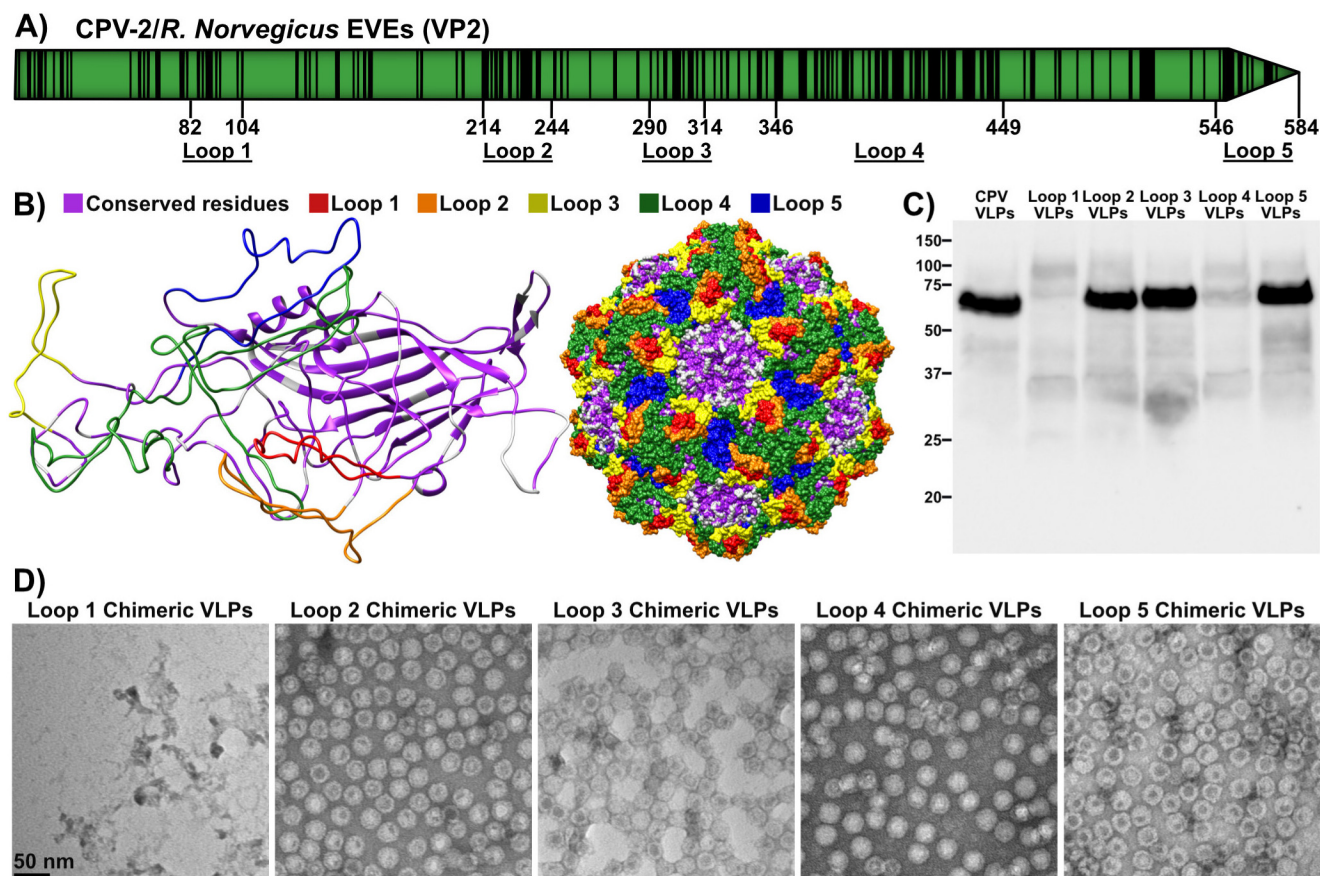
FIG 5 Construction and expression of chimeric VLPs consisting of *R. norvegicus* EVE surface loops incorporated into CPV VP-2. (A) VP2 from the *R. norvegicus* EVE, showing the location of 5 surface loops and their differences from CPV-2. Numbering of residues is from CPV-2. Black bands indicate residues that differ between CPV-2 and the *R. norvegicus* EVE. (B) Diagram of the *R. norvegicus* EVE surface loop/CPV-2 chimeras. *R. norvegicus* EVE surface loops are colored and shown on the CPV-2 VP2 backbone (PDB ID 2CAS) (left) and on the capsid surface (right). (C) Western blot of chimeric VLPs stained with a polyclonal anti-FPV antibody. (D) Electron micrographs of chimeric VLPs.

protoparvoviruses, with a conserved β-barrel structure that made up the core of the assembled capsid and large loops forming most of the exterior surface (Fig. 8B). There was a pocket within the dimple of the capsid, near the 2-fold axis, which was similar to the sialic acid binding sites identified for other parvoviruses (26–29) (Fig. 8C). However, compared to that in the CPV structure, the binding site lacks a conserved glutamic acid residue (Glu396) (Fig. 8D), which is associated with CPV sialic acid binding (29).

DISCUSSION

The initial discovery of integrated sequences related to autonomous parvoviruses in the genomes of various animals was unexpected, as these viruses do not pass through an integrated stage in their life cycles. However, it is now clear that numerous animal genomes contain one or more integrated parvovirus sequences and that most represent viral DNA remnants that became integrated into the genomes of their ancestral hosts hundreds of thousands to millions of years ago (1, 20–22). Whether these integrated sequences benefit the host or play any biological role is still not known. While endogenized sequences of some other viruses have been coopted for host functions or can act to protect against viral infection (30–33), the degenerated state of the inserted parvoviral genes examined here suggests that they are not likely to have functions requiring intact viral proteins or larger peptides. In other systems, integrated viral sequences have functions that are associated with short RNAs or peptide sequences (34, 35), but such functions have not yet been associated with parvovirus integrated sequences.
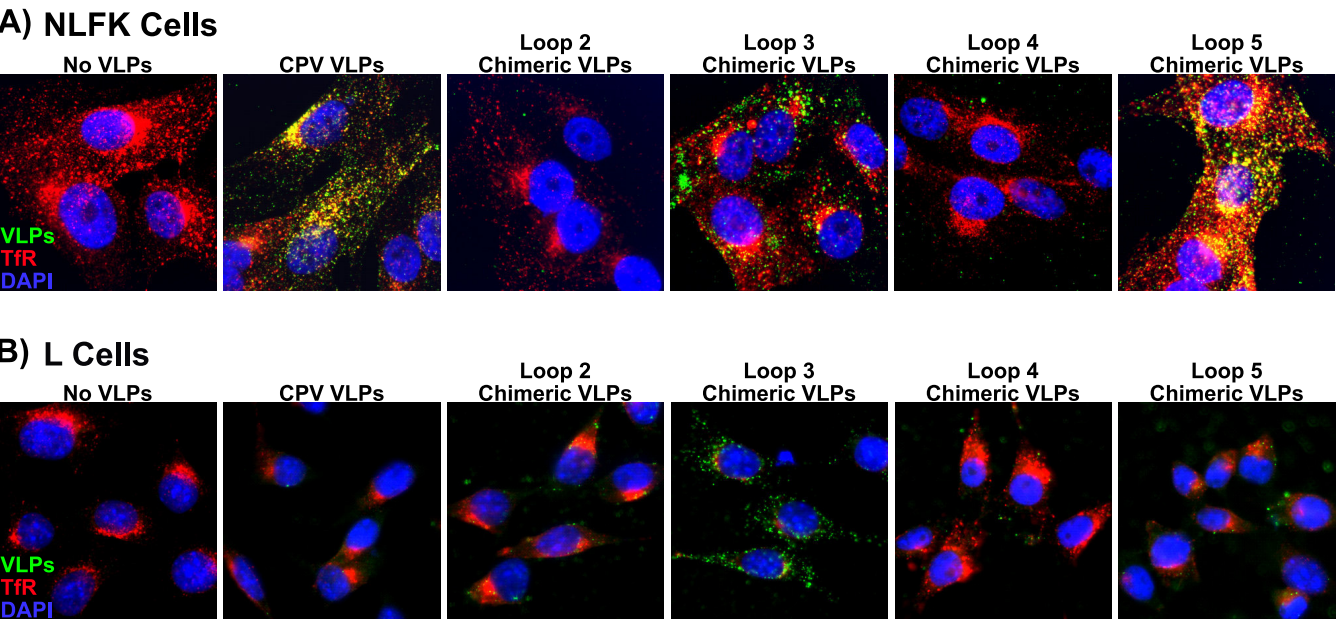
A) NLFK Cells

B) L Cells

FIG 6 Cellular uptake of chimeric *R. norvegicus* EVE/CPV-2 VLPs by NLFK cells (A) or murine L cells (B). Capsids are stained with a polyclonal anti-FPV antibody (green), and the transferrin receptor (TfR) is stained with monoclonal anti-TfR cytoplasmic tail antibody (red).

Here we examined three endogenized major capsid protein gene sequences found in mouse and rat genomes in order to better understand the properties of the likely ancient viruses that they were derived from. The endogenized sequences of the three viruses that we examined were only observed in the genomes of individual species. For
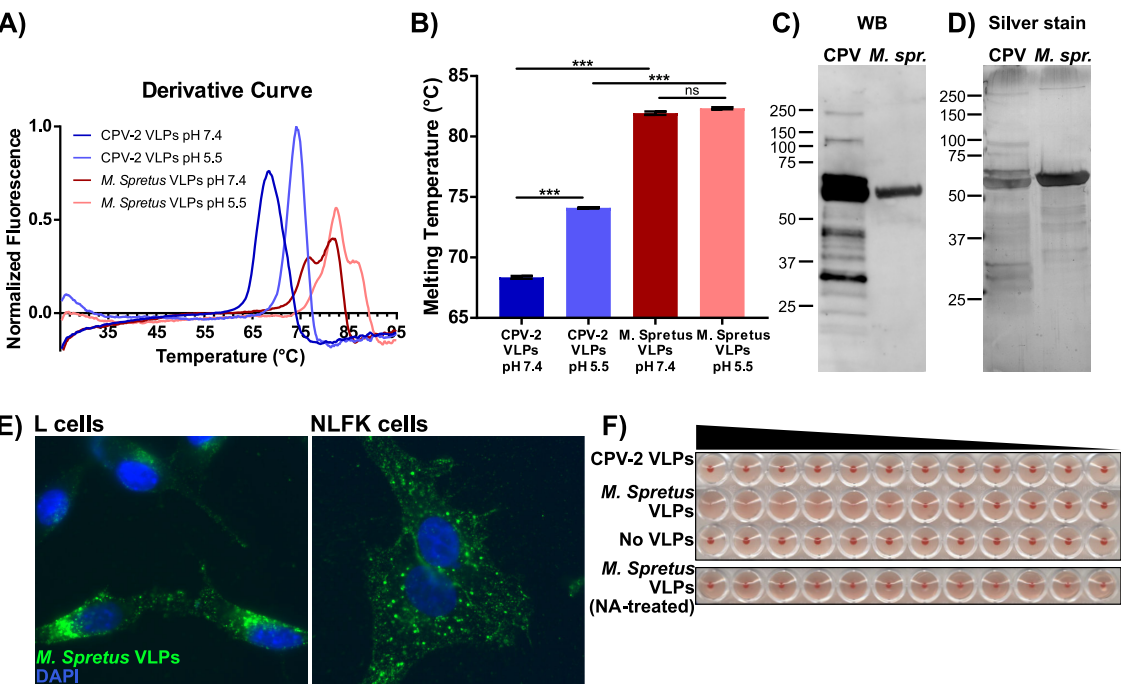

FIG 7 Characterization of assembled *M. spretus* EVE VLPs. (A) Differential scanning fluorometry analysis showing derivative melting curves for CPV-2 VLPs and *M. spretus* EVE VLPs at pH 7.4 and 5.5. (B) Quantification of capsid melting temperature as determined by differential scanning fluorometry. Results of three independent experiments are shown. Western blot (C) and silver stain (D) of purified CPV-2 or *M. spretus* EVE VLPs. The Western blot was stained with a polyclonal anti-FPV antibody. (E) Uptake of fluorescently labeled *M. spretus* EVE VLPs by murine L cells or feline NLFK cells. (F) Hemagglutination assay of CPV-2 VLPs and *M. spretus* EVE VLPs on mouse red blood cells (top) and hemagglutination of *M. spretus* EVE VLPs on mouse red blood cells treated with a broad-spectrum neuraminidase (bottom). Error bars show standard errors. ***, $P < 0.001$.

TABLE 1 Percentages of different modified sialic acids recovered from murine erythrocytes

O-Linked sialic acid type[a]	% abundance
Neu5Ac	60.50
Neu5,9Ac2	16.60
Neu5,7,8/9Ac3	13.90
Neu5,7Ac2	3.40
Neu5,8Ac2	2.90
Neu5Gc	2.70

[a]Neu5,9Ac2, N-acetyl-9-O-acetylneuraminic acid; Neu5,7,8/9Ac3, N-acetyl-7,8/9-di-O-acetylneuraminic acid; Neu5,7Ac2, N-acetyl-7-O-acetylneuraminic acid; Neu5,8Ac2, N-acetyl-8-O-acetylneuraminic acid.

the murid rodents that contained EVEs (*A. sylvaticus* and *M. spretus*), there are a number of whole-genome sequences from closely related animals available, but none contain orthologs of the EVE insertions examined here. Furthermore, the sites corresponding to the *M. spretus* and *A. sylvaticus* EVE insertions in these related species showed no evidence of rearrangement or of integrated sequences, making it unlikely that the insertions were lost from those mice. This indicates that these EVEs were each acquired within the last few million years, since the speciation of those mice. There are few genomic sequences of rats closely related to *R. norvegicus* in databases, and so we examined DNA extracted from the tissues of four closely related rodents for the *R.*
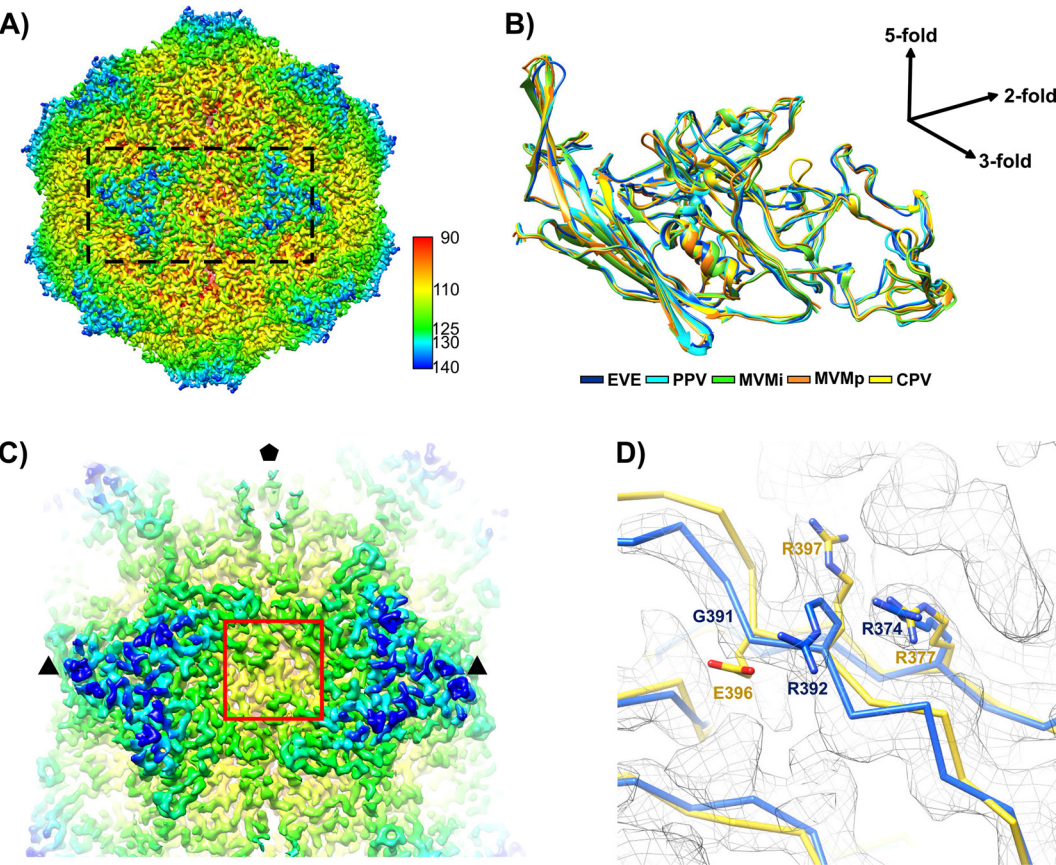

FIG 8 Structure of the *M. spretus* EVE VLP determined using cryo-EM. (A) Structure of the complete capsid was reconstructed to 3.89 Å using the RELION program. The map is icosahedrally averaged and radially colored. (B) Structure of the VP2 build (blue) separately compared to the VP2 crystal structures of the closely related extant viruses, including PPV (PDB ID 1K3V; cyan), as well as another rodent parvovirus minute virus of mice (PDB ID 1MVM, orange; 1Z14, green) and canine parvovirus (PDB ID 2CAS; yellow). (C) Comparison of the defined sialic binding site on the capsid of MVM with the similar structure of the *M. spretus* EVE capsid. The predicted sialic acid binding site is shown in the red rectangle on the capsid of *M. spretus* EVE. (D) The sialic acid binding site of the closely related CPV is compared to its equivalent in *M. spretus* EVE, showing that the binding site lacks the conserved glutamic acid residue.

norvegicus EVE sequence. We did not find that sequence in the DNA of any of the other *Rattus* species examined, indicating that the *R. norvegicus* EVE was likely acquired less than 2 million years ago.

Neither the *R. norvegicus* nor *A. sylvaticus* VP2 proteins assembled into capsids when first expressed in the repaired form that we prepared. The *A. sylvaticus* VP2 sequence contained a relatively large 52-base (17 residue) deletion in the N terminus of the VP2 structure that we filled in with nucleotides from the corresponding region of the *M. spretus* EVE, and incompatible residues within that sequence may have prevented capsid assembly. However, it is also possible that other mutations within the *A. sylvaticus* or *R. norvegicus* EVE sequences or our other repair sequences prevented assembly. Studies with extant parvoviruses show that single point mutations within the major capsid protein can block capsid assembly, especially when they occur within the well-conserved residues that form the interior of the capsid or are near the 5-fold axis (36–39). Expressing many of the less conserved exterior surface loops in place of the equivalent sequences of the CPV capsid allowed the chimeras to assemble (Fig. 4). However, when we simply replaced thirds of the VP2 sequence of CPV with the equivalent repaired sequences from the *R. norvegicus* EVE, all replacements prevented capsids from assembling, showing either that all three regions contain mutations that prevent capsid assembly or that the mismatch between the EVE and CPV sequences creates that effect.

The *R. norvegicus*/CPV chimera containing all 5 EVE surface loops also failed to assemble into capsids, likely because it contained surface loop 1 from the EVE, which prevented capsid assembly when it was used alone to replace the corresponding region in the CPV VP2. The loop 3 chimera altered the uptake of VLPs by feline and murine cells (Fig. 6), suggesting that the loop 3 substitution prevented the chimeric VLPs from binding the feline TfR. Those chimeric VLPs did enter both feline and murine cells, which apparently share a receptor that mediates uptake of these VLPs. Chimeric VLPs with loop 2 or 4 substitutions were not taken up by feline cells. This is unsurprising, as loop 2 has been directly implicated in binding to the feline TfR, and loop 4 substitution, which replaces a large portion of the capsid surface, likely altered the conformation of adjacent TfR binding surface loops.

The *M. spretus* EVE produced capsids directly after being repaired, and these capsids were highly stable at both neutral and endosomal pH. Minor peaks on the DSF curve might be the result of partial unfolding of the capsid or the disassociation of capsid subunits from the whole, either of which would expose hydrophobic residues and result in increased fluorescence. The structure of the *M. spretus* EVE capsids at a resolution of 3.89 Å confirmed that the capsid structure was very similar to those of other rodent and porcine parvoviruses, with some rearrangements of the variable surface loops, as is customarily seen when comparing these viruses (26, 40). *M. spretus* capsids showed uptake patterns by murine cells that were similar to those of extant mouse parvoviruses, with capsids having a perinuclear localization 1 h after uptake (25) (Fig. 7). Those capsids bind Neu5Ac sialic acids, which likely serve as a receptor for cellular uptake and infection, as has been seen for other rodent parvoviruses (40).

In summary, the *M. spretus* EVE sequence assembled into VLPs which bound to sialic acids and entered mouse cells as expected, since the original virus presumably had rodents as the host range. The *R. norvegicus* EVE revealed some of the surface features of the ancient virus capsids. Studying these capsids provides an understanding of how parvoviruses have evolved over thousands or millions of years. It is still unknown whether endogenized parvovirus sequences provide any benefit to their hosts, and future experiments may reveal whether expression of truncated proteins or RNA from these sequences has any effect on the hosts or infecting viruses.

MATERIALS AND METHODS

Genome sequence analysis. The database-integrated genome screening (DIGS) tool (41) (available at http://giffordlabcvr.github.io/DIGS-tool/) was used to identify parvovirus-related sequences among the whole-genome shotgun sequence data of 171 mammal species. A subset of whole-genome shotgun contigs that contained complete or near-complete parvovirus capsid gene sequences and a region of the

TABLE 2 Specimens used as sources of DNA for determining the presence of the rat EVE integrated DNA

Species	No. (sex) of sample sources[a]	FMNH no.[b]
Rattus rattus	2 (M), 2 (F)	222595, 198176, 181086, 179191
Rattus exulans	2 (M), 2 (F)	198761, 168962, 188484, 168964
Rattus everetti	2 (M), 2 (F)	198872, 198796, 196061, 191246
Rattus tanezumi	2 (M), 2 (F)	198768, 198797, 194747, 178427

[a]M, male; F, female.
[b]FMNH, Field Museum of Natural History.

host genomic sequence were selected for further analysis. A combination of automated procedures (MUSCLE [42, 43] and BLAST [44]) and manual adjustments were used to align parvovirus EVEs to the most closely related extant virus sequences among a set of representative parvovirus genomes, allowing us to infer the likely ancestral open reading frames (ORFs). The original alignments were also used to construct phylogenies using maximum likelihood (ML) as implemented in RAxML (45). The parameters for phylogenetic analysis were selected using ProtTest (46).

Analysis of the viral sequences and reconstruction of open reading frames. Three endogenized viral sequences with more or less full-length capsid protein genes derived from the genomes of *R. norvegicus*, *M. spretus*, and *A. sylvaticus* were chosen for protein expression. These sequences had relatively complete open reading frames and homology to contemporary infectious viruses. To repair the parvovirus EVEs, VP2 sequences were aligned with the sequences of a group of the most closely related contemporary infectious viruses, and the open reading frames were modified to remove stop codons and predicted insertions and deletions, recreating the open reading frame (diagrammed in Fig. 1). To repair likely deletions, we introduced an amino acid or sequence that was most similar to the consensus of the closely related contemporary viral sequences (see Data Set S3 in the supplemental material).

Identification of EVE sequences in DNA species related to *R. norvegicus*. To obtain more information about the possible presence of the sequence within the *R. norvegicus* genome, we obtained DNA from tissue samples from *R. norvegicus* laboratory rats, as well as from bio-banked samples of *R. rattus*, *R. tanezumi*, *R. exulans*, and *R. everetti* from the Field Museum of Natural History, Chicago (Table 2). The *R. norvegicus* rats were not in our possession, and tissues were obtained from animals that had been euthanized in the course of other studies. DNA was extracted from tissues using a Tissue DNA extraction kit (Omega Bio-tek). Primers that recognized the EVE sequence were used to perform PCR (5′-AATGTATTGGTCGTATGCTTCGTCGTG-3′ and 5′-CCCAACTTGGTCCGAAATC-3′), with a total of 100 μg of DNA and 35 cycles of PCR with Q5 High Fidelity DNA polymerase (NEB). Primers that recognized a conserved region in retinol binding protein 3 (IRBP) were used as a positive control for PCR (5′-TCTCA GCTTCTGGAGGTC-3′ and 5′-CTGCTGGCCCAGATACAGAG-3′).

Cells. Human embryonic kidney 293 (HEK293) cells, Sf9 (*Spodoptera frugiperda*) cells (Invitrogen), and High Five (*Trichoplusia ni*) cells (clone BTI-TN-551-4; Boyce Thompson Institute) were used in this study. HEK293 cells were grown in Dulbecco modified Eagle's medium (DMEM) with 10% fetal calf serum (FCS) at 37°C, Sf9 cells were grown in Grace's supplemented insect medium (Gibco) with 10% FCS at 28°C, and High Five cells were grown in Express5 serum-free medium (Gibco) at 28°C. Norden laboratory feline kidney (NLFK) cells (Norden Laboratories) and L cells (a murine fibroblast cell line obtained from the laboratory of John Parker, Cornell University) were also used in this study. NLFK cells were grown in 1:1 McCoy's/Lebovitz medium (Lonza) with 5% FCS at 37°C, and L cells were grown in Eagle's minimal essential medium (EMEM) (Lonza) with 10% FCS at 37°C.

Expression of endogenous viral element capsid proteins. The complete, repaired VP2 sequences were codon optimized for mammalian cell expression and synthesized along with the 12 bp preceding the start codon of canine parvovirus VP2 to favor protein expression. The *R. norvegicus* EVE sequence was then cloned into the pCDNA3.1(−) plasmid vector under the control of the cytomegalovirus (CMV) immediate early promoter for expression in HEK293 cells. The VP2 genes for all the EVEs were also cloned into the pFastBac1 plasmid (Invitrogen) under the control of the polyhedron promoter for baculovirus expression in High Five cells. Baculoviruses were prepared by transfection of bacmids into Sf9 cells, and then proteins prepared by inoculation of the viruses into High Five insect cells, as described previously (36).

For VP2 capsid protein purification, HEK293 cells were collected 3 days after transfection with plasmids or 4 days after infection with baculovirus. Supernatants from transfected HEK293 cells were pelleted at $100,000 \times g$ for 2 h. Virus-like particles (VLPs) were purified from infected High Five cells as described previously (36). The resulting fractions were examined via negative stain electron microscopy with 2% uranyl acetate to determine whether capsids were present.

Construction of *R. norvegicus* EVE/CPV-2 chimeric VP2 sequences. Some of the VP2 sequences did not assemble into capsids when expressed from plasmids after transfection into HEK293 cells or after infection of High Five cells with baculovirus expressing VP2 sequences. To examine the nature of any assembly defects in the *R. norvegicus* EVE VP2 gene sequence and to allow definition of the structures of the protein domains, we prepared recombinants that combined the *R. norvegicus* EVE VP2 sequence with that of the CPV-2 VP2 gene. The first set of chimeras were constructed by digesting the *R. norvegicus* EVE and CPV-2 VP2 genes with BglII and PhoI restriction enzymes (NEB), dividing the genes roughly into thirds, and ligating the sequences together in all 6 possible combinations, as shown in Fig. 4. The second set of chimeras was created by cloning surface loops from the *R. norvegicus* EVE into the CPV-2 VP2

backbone, as depicted in Fig. 5. Five loops were individually cloned into the backbone (B-C loop [loop 1, replacing CPV VP2 residues 82 to 104], E-F loop [loop 2, residues 214 to 244], G-H loop [loop 3, residues 290 to 314], G-H loop [loop 4, residues 346 to 449], and the C-terminal sequence [loop 5, residues 546 to 584]), as well as a chimera that contained all 5 surface loops. Capsid assembly for each set of chimeras was determined by purifying capsids as described above and performing negative stain electron microscopy with 2% uranyl acetate.

Examination of assembled capsid properties. The *M. spretus* EVE VLPs and wild-type CPV-2 VLPs were tested for stability using differential scanning fluorometry. For this, 2.5 μg of VLPs in either phosphate-buffered saline (PBS; pH 7.4) or 0.1 M citric acid (pH 5.5) were combined with 2.5 μl of 1% SYPRO orange dye (Life Technologies) in a total volume of 25 μl and heated from 25°C to 95°C at a rate of 1°C/min using a StepOnePlus real-time PCR thermocycler (Applied Biosystems). Melting temperature was determined by taking the first derivative of the resulting curve and finding its maximum. Results from three independent experiments were analyzed via analysis of variance (ANOVA) using GraphPad Prism (v7.04; GraphPad Software, Inc.).

Purified CPV-2 and *M. spretus* EVE VLPs were separated on a 10% SDS-PAGE gel and either silver stained using a Pierce Silver Stain kit (Thermo) according to the manufacturer's instructions or transferred onto a nitrocellulose membrane and stained with a rabbit polyclonal anti-FPV primary antibody followed by a goat anti-rabbit horseradish peroxidase (HRP) secondary antibody and ECL Western Blotting substrate (Pierce).

Capsids were also tested for hemagglutination (HA) with mouse erythrocytes. Capsids were serially diluted 1:1 in 25 μl of bis-Tris-buffered saline (BTBS), pH 6.2. Then, 50 μl of 0.5% (vol/vol) mouse erythrocytes in BTBS were added, and mixtures were incubated at 4°C overnight. Neuraminidase (NA) treatment of erythrocytes involved overnight incubation at 37°C with 40 U of α-2,3,6,8,9 neuraminidase A (NEB) in PBS, pH 7.4. The sialic acid forms on mouse erythrocytes were analyzed after release using 2 M acetic acid at 80°C for 3 h and labeling with 1,2-diamino-4,5-methylenedioxybenzene (DMB; Sigma-Aldrich) for 2.5 h at 50°C (47). High-pressure liquid chromatography (HPLC) analysis was performed using a Dionex UltiMate 3000 system with an Acclaim C_{18} column (Thermo Fisher) under isocratic elution in 7% methanol, 7% acetonitrile, and 86% water. Sialic acid standards were bovine submaxillary mucin and commercial standards for Neu5Ac and *N*-glycolylneuraminic acid (Neu5Gc; Sigma-Aldrich).

Cell binding and uptake. *M. spretus* EVE VLPs were fluorescently labeled with Alexa 488 using Alexa Fluor 488 5-SDP ester (Life Technologies) according to the manufacturer's protocol. Twenty micrograms of capsids was added to mouse L cells seeded at 1×10^5 cells/cm² on glass coverslips, and the cells were incubated for 1 h at 37°C. Cells were then fixed in 4% paraformaldehyde for 10 min, mounted on slides using ProLong Gold antifade mountant with DAPI (4',6-diamidino-2-phenylindole; Life Technologies), and imaged using a Nikon TE300 epifluorescence microscope. Cell binding and uptake of loop 2 to loop 5 chimeric VLPs were determined by incubating 20 μg/ml capsids with NLFK or L cells seeded at 2×10^4 cells/cm² on coverslips. Cells were fixed with 4% paraformaldehyde for 10 min and then stained with a rabbit anti-FPV polyclonal antibody and a mouse anti-TfR monoclonal antibody (QB213080; Life Technologies), followed by staining with goat anti-rabbit Alexa 488 and/or goat anti-mouse Alexa 594 secondary antibodies. Cells were mounted on slides and imaged in the same way as for the *M. spretus* VLPs.

Cryo-EM. The *M. spretus* EVE sample was applied to glow-discharged Quantifoil EM grids (Electron Microscopy Sciences), which were blotted and vitrified in liquid ethane using an FEI Vitrobot Mark IV. Images were acquired on the Titan Krios G3 microscope (Thermo Fisher Scientific) operated with an accelerating voltage of 300 kV. An "Atlas" image was assembled from micrographs taken at ×165 magnification in linear mode on the Falcon 3ec direct electron detector, and suitable areas were selected for data collection. Automated data collection was setup using FEI's EPU software. Images were collected on the Falcon 3ec direct electron detector in counting mode using a nominal magnification of ×59,000, resulting in a calibrated pixel size of 1.136 Å. The microscope was operated with a 70-μm condenser aperture and a 100-μm objective aperture. Four nonoverlapping exposures per 2-μm-diameter hole in a Quantifoil R 2/1 were acquired with the beam in parallel mode. Total dose per exposure was set at 45 e$^-$/Å².

We collected a total of 2,082 images and selected 6,432 particles for the final reconstruction. We used the structure of PPV as a model for building into the map, as the crystal structure of PPV VP2 (PDB identifier [ID] 1K3V) fit into the cryo-EM map with a correlation coefficient of 0.722 as measured by the Chimera Fit in Map program (48). We substituted the sequence of PPV VP2 with the repaired *M. spretus* VP2 that had been expressed to prepare the VLPs using the program COOT (49). The build was further refined using the PHENIX real space refine program (50). The final build fit into the cryo-EM map with a correlation coefficient of 0.8781 and was validated using the Molprobity program (51, 52).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/JVI.01542-18.

SUPPLEMENTAL FILE 1, XLSX file, 0.02 MB.

ACKNOWLEDGMENTS

REFERENCES

1. Smith RH, Hallwirth CV, Westerman M, Hetherington NA, Tseng Y-S, Cecchini S, Virag T, Ziegler M-L, Rogozin IB, Koonin EV, Agbandje-McKenna M, Kotin RM, Alexander IE. 2016. Germline viral "fossils" guide *in silico* reconstruction of a mid-Cenozoic era marsupial adeno-associated virus. Sci Rep 6:28965. https://doi.org/10.1038/srep28965.

2. Im DS, Muzyczka N. 1990. The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity. Cell 61:447–457.

3. Samulski RJ, Zhu X, Xiao X, Brook JD, Housman DE, Epstein N, Hunter LA. 1991. Targeted integration of adeno-associated virus (AAV) into human chromosome 19. EMBO J 10:3941–3950. https://doi.org/10.1002/j.1460-2075.1991.tb04964.x.

4. Linden RM, Winocour E, Berns KI. 1996. The recombination signals for adeno-associated virus site-specific integration. Proc Natl Acad Sci U S A 93:7966–7972.

5. Cotmore SF, Agbandje-McKenna M, Chiorini JA, Mukha DV, Pintel DJ, Qiu J, Soderlund-Venermo M, Tattersall P, Tijssen P, Gatherer D, Davison AJ. 2014. The family *Parvoviridae*. Arch Virol 159:1239–1247. https://doi.org/10.1007/s00705-013-1914-1.

6. Steinel A, Parrish CR, Bloom ME, Truyen U. 2001. Parvovirus infections in wild carnivores. J Wildl Dis 37:594–607. https://doi.org/10.7589/0090-3558-37.3.594.

7. Mendelson E, Trempe JP, Carter BJ. 1986. Identification of the trans-acting Rep proteins of adeno-associated virus by antibodies to a synthetic oligopeptide. J Virol 60:823–832.

8. Miller CL, Pintel DJ. 2002. Interaction between parvovirus NS2 protein and nuclear export factor Crm1 is important for viral egress from the nucleus of murine cells. J Virol 76:3257–3266. https://doi.org/10.1128/JVI.76.7.3257-3266.2002.

9. Redemann BE, Mendelson E, Carter BJ. 1989. Adeno-associated virus rep protein synthesis during productive infection. J Virol 63:873–882.

10. Wang D, Yuan W, Davis I, Parrish CR. 1998. Nonstructural protein-2 and the replication of canine parvovirus. Virology 240:273–281. https://doi.org/10.1006/viro.1997.8946.

11. Hernando E, Llamas-Saiz AL, Foces-Foces C, McKenna R, Portman I, Agbandje-McKenna M, Almendral JM. 2000. Biochemical and physical characterization of parvovirus minute virus of mice virus-like particles. Virology 267:299–309. https://doi.org/10.1006/viro.1999.0123.

12. López de Turiso JA, Cortés E, Martínez C, Ruiz de Ybáñez R, Simarro I, Vela C, Casal I. 1992. Recombinant vaccine for canine parvovirus in dogs. J Virol 66:2748–2753.

13. Ruffing M, Zentgraf H, Kleinschmidt JA. 1992. Assembly of viruslike particles by recombinant structural proteins of adeno-associated virus type 2 in insect cells. J Virol 66:6922–6930.

14. Girod A, Wobus CE, Zádori Z, Ried M, Leike K, Tijssen P, Kleinschmidt JA, Hallek M. 2002. The VP1 capsid protein of adeno-associated virus type 2 is carrying a phospholipase A2 domain required for virus infectivity. J Gen Virol 83:973–978. https://doi.org/10.1099/0022-1317-83-5-973.

15. Vihinen-Ranta M, Wang D, Weichert WS, Parrish CR. 2002. The VP1 N-terminal sequence of canine parvovirus affects nuclear transport of capsids and efficient cell infection. J Virol 76:1884–1891.

16. Ahn JK, Pitluk ZW, Ward DC. 1992. The GC box and TATA transcription control elements in the P38 promoter of the minute virus of mice are necessary and sufficient for transactivation by the nonstructural protein NS1. J Virol 66:3776–3783.

17. Cotmore SF, Tattersall P. 1994. An asymmetric nucleotide in the parvoviral 3′ hairpin directs segregation of a single active origin of DNA replication. EMBO J 13:4145–4152.

18. Krauskopf A, Resnekov O, Aloni Y. 1990. A *cis* downstream element participates in regulation of *in vitro* transcription initiation from the P38 promoter of minute virus of mice. J Virol 64:354–360.

19. Tyson JJ, Chen KC, Lederman M, Bates RC. 1990. Analysis of the kinetic hairpin transfer model for parvoviral DNA replication. J Theor Biol 144:155–169.

20. Arriagada G, Gifford RJ. 2014. Parvovirus-derived endogenous viral ele-

ments in two South American rodent genomes. J Virol 88:12158–12162. https://doi.org/10.1128/JVI.01173-14.

21. Katzourakis A, Gifford RJ. 2010. Endogenous viral elements in animal genomes. PLoS Genet 6:e1001191. https://doi.org/10.1371/journal.pgen.1001191.

22. Liu H, Fu Y, Xie J, Cheng J, Ghabrial SA, Li G, Peng Y, Yi X, Jiang D. 2011. Widespread endogenization of densoviruses and parvoviruses in animal and human genomes. J Virol 85:9863–9876. https://doi.org/10.1128/JVI.00828-11.

23. Rowe KC, Aplin KP, Baverstock PR, Moritz C. 2011. Recent and rapid speciation with limited morphological disparity in the genus *Rattus*. Syst Biol 60:188–203. https://doi.org/10.1093/sysbio/syq092.

24. Tsao J, Chapman MS, Agbandje M, Keller W, Smith K, Wu H, Luo M, Smith TJ, Rossmann MG, Compans RW. 1991. The three-dimensional structure of canine parvovirus and its functional implications. Science 251:1456–1464.

25. Parker JS, Murphy WJ, Wang D, O'Brien SJ, Parrish CR. 2001. Canine and feline parvoviruses can use human or feline transferrin receptors to bind, enter, and infect cells. J Virol 75:3896–3902. https://doi.org/10.1128/JVI.75.8.3896-3902.2001.

26. Pittman N, Misseldine A, Geilen L, Halder S, Smith JK, Kurian J, Chipman P, Janssen M, Mckenna R, Baker TS, D'Abramo A, Cotmore S, Tattersall P, Agbandje-McKenna M. 2017. Atomic resolution structure of the onco-lytic parvovirus LuIII by electron microscopy and 3D image reconstruction. Viruses 9:E321. https://doi.org/10.3390/v9110321.

27. Nam H-J, Gurda-Whitaker B, Gan WY, Ilaria S, McKenna R, Mehta P, Alvarez RA, Agbandje-McKenna M. 2006. Identification of the sialic acid structures recognized by minute virus of mice and the role of binding affinity in virulence adaptation. J Biol Chem 281:25670–25677. https://doi.org/10.1074/jbc.M604421200.

28. López-Bueno A, Rubio M-P, Bryant N, McKenna R, Agbandje-McKenna M, Almendral JM. 2006. Host-selected amino acid changes at the sialic acid binding pocket of the parvovirus capsid modulate cell binding affinity and determine virulence. J Virol 80:1563–1573. https://doi.org/10.1128/JVI.80.3.1563-1573.2006.

29. Tresnan DB, Southard L, Weichert W, Sgro JY, Parrish CR. 1995. Analysis of the cell and erythrocyte binding activities of the dimple and canyon regions of the canine parvovirus capsid. Virology 211:123–132. https://doi.org/10.1006/viro.1995.1385.

30. Bézier A, Herbinière J, Lanzrein B, Drezen J-M. 2009. Polydnavirus hidden face: the genes producing virus particles of parasitic wasps. J Invertebr Pathol 101:194–203. https://doi.org/10.1016/j.jip.2009.04.006.

31. Ito J, Watanabe S, Hiratsuka T, Kuse K, Odahara Y, Ochi H, Kawamura M, Nishigaki K. 2013. Refrex-1, a soluble restriction factor against feline endogenous and exogenous retroviruses. J Virol 87:12029–12040. https://doi.org/10.1128/JVI.01267-13.

32. Lavialle C, Cornelis G, Dupressoir A, Esnault C, Heidmann O, Vernochet C, Heidmann T. 2013. Paleovirology of "syncytins", retroviral env genes exapted for a role in placentation. Philos Trans R Soc Lond B Biol Sci 368:20120507. https://doi.org/10.1098/rstb.2012.0507.

33. Malfavon-Borja R, Feschotte C. 2015. Fighting fire with fire: endogenous retrovirus envelopes as restriction factors. J Virol 89:4047–4050. https://doi.org/10.1128/JVI.03653-14.

34. Yang N, Kazazian HH. 2006. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. Nat Struct Mol Biol 13:763–771. https://doi.org/10.1038/nsmb1141.

35. Belyi VA, Levine AJ, Skalka AM. 2010. Unexpected inheritance: multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes. PLoS Pathog 6:e1001030. https://doi.org/10.1371/journal.ppat.1001030.

36. Callaway HM, Feng KH, Lee DW, Allison AB, Pinard M, McKenna R, Agbandje-McKenna M, Hafenstein S, Parrish CR. 2017. Parvovirus capsid structures required for infection: mutations controlling receptor recognition and protease cleavages. J Virol 91:e01871-16. https://doi.org/10.1128/JVI.01871-16.

37. Kawase M, Momoeda M, Young NS, Kajigaya S. 1995. Modest truncation

of the major capsid protein abrogates B19 parvovirus capsid formation. J Virol 69:6567–6571.

38. Lombardo E, Ramírez JC, Agbandje-McKenna M, Almendral JM. 2000. A beta-stranded motif drives capsid protein oligomers of the parvovirus minute virus of mice into the nucleus for viral assembly. J Virol 74: 3804–3814.

39. Yuan W, Parrish CR. 2001. Canine parvovirus capsid assembly and differences in mammalian and insect cells. Virology 279:546–557. https://doi.org/10.1006/viro.2000.0734.

40. Huang L-Y, Halder S, Agbandje-McKenna M. 2014. Parvovirus glycan interactions. Curr Opin Virol 7:108–118. https://doi.org/10.1016/j.coviro.2014.05.007.

41. Zhu H, Dennis T, Hughes J, Gifford RJ. 2018. Database-integrated genome screening (DIGS): exploring genomes heuristically using sequence similarity search tools and a relational database. bioRxiv https://doi.org/10.1101/246835.

42. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.

43. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113. https://doi.org/10.1186/1471-2105-5-113.

44. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

45. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033.

46. Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27: 1164–1165. https://doi.org/10.1093/bioinformatics/btr088.

47. Varki A, Diaz S. 1984. The release and purification of sialic acids from glycoconjugates: methods to minimize the loss and migration of O-acetyl groups. Anal Biochem 137:236–247.

48. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera–a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612. https://doi.org/10.1002/jcc.20084.

49. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66:486–501. https://doi.org/10.1107/S0907444910007493.

50. Afonine PV, Poon BK, Read RJ, Sobolev OV, Terwilliger TC, Urzhumtsev A, Adams PD. 2018. Real-space refinement in PHENIX for cryo-EM and crystallography. Acta Crystallogr D Struct Biol 74:531–544. https://doi.org/10.1107/S2059798318006551.

51. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66:12–21. https://doi.org/10.1107/S0907444909042073.

52. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, Verma V, Keedy DA, Hintze BJ, Chen VB, Jain S, Lewis SM, Arendall WB, Snoeyink J, Adams PD, Lovell SC, Richardson JS, Richardson DC. 2018. MolProbity: more and better reference data for improved all-atom structure validation. Protein Sci 27:293–315. https://doi.org/10.1002/pro.3330.